AMENDMENT OF THE SPECIFICATION

Please replace the second full paragraph on page 2 with the following amended paragraph.

Thus, there is a need for a computer implemented statistical modeling program that is flexible enough to analyze data sets comprising a plurality of independent variables, but which provides a meaningful mathematical description of the data set. For example, it would be desirable to have the statistical modeling analysis describe the data using a minimum number of terms, so that the significance of each independent variable can be evaluated in a meaningful manner. There is also a need for software that can automatically approximate values for missing data. It would also be beneficial to have a statistical modeling method that provides a series of increasingly complex equations, so that a user can apply the data set to real world problems, and evaluate the models provided by the analysis in light of known physical parameters.

Please replace the first full paragraph on page 7 with the following amended paragraph.

In an embodiment, the set of functions used to fit the independent variables to the dependent variable and residuals of the dependent variable may be the same at each fitting step (*e.g*, F_{Sn} , = F_{Sn-1} = F_{S3} = F_{S2} = F_{S1}), thereby simplifying program step selection. Alternatively, the set used to fit a less important variable may be larger that sets used to fit more important independent variables (*e.g.*, F_{Sn} , > F_{Sn-1}, > F_{S3} > F_{S2} > F_{S1}) since the functions that explain a less important variable (*e.g.*, x_3 and x_2) in relation to $y - \hat{y}_1$ may not be in the first set of functions (F_{S1}) required to explain the most important variable (*e.*g., x_1). Alternatively, the set used to fit a more important variable may be larger than ~~that~~ sets used to fit less important independent variables (*e.g.*, $F_{S3} < F_{S2} < F_{S1}$), as the function used to define the most important variable (*e.*g., x_1) is not needed to define less important variables.

Please replace the second full paragraph on page 16 with the following amended paragraph.

The present invention is distinct from other techniques described ~~describe~~ for automated or semi-automated statistical modeling. The early application of residual analysis as a means for

statistical modeling using a computer required extensive human interpretation of the data during the statistical modeling process (*see e.g.,* Ingels, R., Chem. Engineering, August 11, 1980, pp. 145-156) and thus, was not practical or even workable for large data sets. Other applications for computerized statistical analysis have been developed for analysis of predetermined variables, such as how a manifest variable impacts on a latent variable (U.S. Patent No. 6,192,319) or the use of residual analysis to analyze clustering of data for finding underlying patterns in the data set (U.S. Patent No. 6,026,397). Other patents relate to automatic report generation, but do not provide a mathematical analysis (U.S. Patent No. 6,055,541). Thus, the present invention fills a need in the field of providing a mathematical description of a previously unprocessed data set that can be used to analyze the data in terms of the most important variables.

Please replace the third full paragraph of Example 2 on page 25 with the following amended paragraph.

For this data set Equation A is the chosen model. Yet for many purposes ~~purpose~~, Equation B or C may suffice. As used herein, LOG () is the natural logarithm (base e). EXP (X) is e raised to the X power (approximately 2.7183**X). WILDLOG () is the natural logarithm (base e) and is defined to be 0.0 if the argument is less than 1.0. This eliminates the spikes occasionally seen in LOG() caused by negative numbers.